

Figure 1: Genome Threader output for INSP163 polypeptide sequence (SEQ ID NO: 30)

## Genome Threader results - Energy Scores

Num	PDB Code	Norm Align Score	Raw Align Score	%IDs Aligned	%Struc Aligned	%Query Seq Aligned	Pair-wise Energy	Solvation Energy	Net Score	%Confidence for Query	From Pos for Query	To Pos for Query	From Pos for Target	To Pos for Target	Alignment Length
1	1cr3AA00 (align)	67.63	96	19.8	75.0	37.7	-155.89	-4.34	0.983	96	187	300	31	129	116
2	1c28AA00 (align)	65.86	94	17.4	72.8	37.7	-105.30	-2.65	0.983	96	187	300	31	121	115
3	1c28cc00 (align)	51.74	79	20.2	70.6	37.7	-9.01	-3.13	0.977	92	187	300	28	99	114
4	1c288900 (align)	45.38	73	18.2	29.7	10.9	-6.72	-2.55	0.971	88	187	219	25	57	33
5	1c39AA00 (align)	42.14	71	20.3	82.4	39.1	-94.33	1.68	0.965	84	182	299	25	146	128
6	2mf1AA00 (align)	39.14	68	17.3	81.8	39.1	-69.55	-1.51	0.961	81	182	299	26	146	129
7	1c39AA00 (align)	34.25	63	19.0	77.8	38.7	-43.14	3.51	0.957	69	183	299	30	141	121
8	1d28AA00 (align)	32.56	61	16.8	75.4	39.4	-12.75	4.41	0.943	72	183	301	34	134	119
9	1d66AA00 (align)	31.11	60	15.3	80.5	39.7	-7.91	-3.18	0.933	68	182	301	30	149	131

1 tgagccgcct cgggacggag ccattgcggcg ctgggacctg gccgcggtcg tggctctcct  
m r r w a w a a v v v l  
INSP163-CP1

61 cgggccgcag ctctgtctcc tcggggggcgt cggggcccgg cgggaggcac agaggacgca  
l g p q l v l l g g v g a r r e a q r t

121 gcagcctggc cagcgcgagc atccccccaa cgccaccgcc agcgcgctcct cccgcgaggg  
q q p g q r a d p p n a t a s a s s r e

181 gctgcccagc gcccacaagc catcccaggc ctcaggacct gagttctcgc acgccacat  
g l p e a p k p s q a s g p e f s d a h

241 gacatggctg aactttgtcc ggcgccggga cgacggcgcc ttaaggaagc ggtgcggaag  
m t w l n f v r r p d d g a l r k r c g

301 cagggacaag aagccgcggg atctcttcgg tccccagga cctccagggt cagaaagtgc  
s r d k k p r d l f g p p g p p g a e v

361 cgcgagact ctgcttcacg agtttcagga gctgctgaaa gaggccacgg agcgccggtt  
t a e t l l h e f q e l l k e a t e r r

421 ctcagggtt ctggaccgcg tgetgcccga gggggcgggc ctgcggttgg tgggcgaggg  
f s g l l d p l l p q g a g l r l v g e

481 ctttcaactgc cggctgcagg gtccccgcgg ggtggacaag cggacgctgg tggagctgca  
a f h c r l q g p r r v d k r t l v e l

541 tggtttccag gctcctgctg cccaagggtc cttcctgcca ggctccggtc tgagccctggc  
h g f q a p a a q g a f l r g s g l s l

601 ctcgggtcgg ttcacggccc ccgtgtccgg catcttccag ttctctgcca gtctgcacgt  
a s g r f t a p v s g i f q f s a s l h

661 ggaccacagt gagctgcagg gcaaggcccg gctgcggggc cgggacgtgg tgtgtgttct  
v d h s e l q g k a r l r a r d v v c v

721 catctgtatt gagtcctgtg gccagcgcca cacgtgcttg gaggccgtct caggcctgga  
l i c i e s l c q r h t c l e a v s g l

781 gagcaacagc agggctcttca cgctacaggt gcaggggctg ctgcagctgc aggctggaca  
e s n s r v f t l q v q g l l q l q a g

841 gtacgcttct gtgtttgtgg acaatggctc cggggccgct ctcaccatcc aggcgggctc  
q y a s v f v d n g s g a v l t i q a g

901 cagcttctcc gggctgctcc tgggcacgt  
s s f s g l l l g t  
INSP163-CP2

Figure 3

**A) INSP163-A**

```

1   PDDGALRKRC  GSRDKKPRDL  FGPPGPPGAE  VTAETLLHEF  QELLKEATER
51  RFSGLLDPLL  PQGAGLRLVG  EAFHCRLQGP  RRVDKRTLVE  LHGFQAPAAQ
101 GAFLRGSGLS  LASGRFTAPV  SGIFQFSASL  HVDHSELQ GK  ARLRARDVVC
151 VLICIESLCQ  RHTCLEAVSG  LESNSRVFTL  QVQGLLQLQA  GQYASVFVDN
201 GSGAVLTIQA  GSSFSGLLLG  T

```

**B) INSP163-B**

```

1   KRCGSRDKKP  RDLFGPPGPP  GAEVTAETLL  HEFQELLKEA  TERRFSGLLD
51  PLLPQGAGLR  LVGEAFHCRL  QGPRRVDKRT  LVELHGFQAP  AAQGAFLRGS
101 GLSLASGRFT  APVSGIFQFS  ASLHVDHSEL  QGKARLRARD  VVCVLICIES
151 LCQRHTCLEA  VSGLESNSRV  FTLQVQGLLQ  LQAGQYASVF  VDNGSGAVLT
201 IQAGSSFSGL  LLGT

```

**C) INSP163-C**

```

1   CGSRDKKPRD  LFGPPGPPGA  EVTAETLLHE  FQELLKEATE  RRFSGLLDPL
51  LPQGAGLRLV  GEAFHCRLQG  PRRVDKRTL V  ELHGFQAPAA  QGAFLRGSGL
101 SLASGRFTAP  VSGIFQFSAS  LHVDHSELQG  KARLRARDVV  CVLICIESLC
151 QRHTCLEAVS  GLESNSRVFT  LQVQGLLQLQ  AGQYASVFVD  NGSGAVLTIQ
201 AGSSFSGLLL  GT

```

**D) INSP163-D**

```

1   FSGLLDPLLP  QGAGLRLVGE  AFHCRLQGPR  RVDKRTLVEL  HGFQAPAAQG
51  AFLRGSGLSL  ASGRFTAPVS  GIFQFSASLH  VDHSELQGKA  RLRARDVVCV
101 LICIESLCQR  HTCLEAVSGL  ESNSRVFTLQ  VQGLLQLQAG  QYASVFVDNG
151 SGAVLTIQAG  SSFSGLLLG

```

**E) INSP163-E**

```

1   VDKRTLVELH  GFQAPAAQGA  FLRGSGLSLA  SGRFTAPVSG  IFQFSASLHV
51  DHSELQGKAR  LRARDVVCVL  ICIESLCQRH  TCLEAVSGLE  SNSRVFTLQV
101 QGLLQLQAGQ  YASVFVDNGS  GAVLTIQAGS  SFSGLLLGT

```

**F) INSP163-F**

```

1   TLVELHGFQA  PAAQGAFLRG  SGLSLASGRF  TAPVSGIFQF  SASLHVDHSE
51  LQGKARLRAR  DVVCVLICIE  SLCQRHTCLE  AVSGLESNSR  VFTLQVQGLL
101 QLQAGQYASV  FVDNGSGAVL  TIQAGSSFSG  LLLGT

```

Figure 4

